CptS 591: Elements of Network Science

Semester Project

Assefaw Gebremedhin
School of Electrical Engineering and Computer Science
assefaw.gebremedhin@wsu.edu
http://www.eecs.wsu.edu/~assefaw

Topics

Covered:

- Introduction
- Graph theory refresher
- Intro to igraph (2)
- Basic network properties (2)
- Random graphs (2)
- Spectral analysis (2)
- Centrality (2)
- Link analysis: PageRank
- Link analysis: Hubs and Authorities
- Network visualization

To come:

- Similarity
- Community identification (clustering)
- Graph similarity with (un)known node correspondence
- (Graph embedding)
- Signed networks
- Cascading behaviors
- Influence maximization
- Epidemic models
- Time varying networks

Project

- Constitutes 40% of final grade
- Team of 2 or 3
- Topic and scope: largely up to you as long as the project
  - Clearly falls within the realm of the course
  - Is likely to be doable within the remainder of the semester
  - Is reasonably novel and interesting (e.g. could form a basis for further research)

Project: four basic types

- *Analysis* of an interesting dataset using existing implementations of algorithms, models and measures.
- *Implementation* of a new algorithm, model or measure and evaluating its performance on a collection of datasets.
- A *theoretical* project that considers an algorithm, a model or a measure and derives rigorous results about it.
- A *critical survey* of a specific topic in an area of the course, going in some depth and offering a fresh perspective.

A project may also (and is encouraged to) be a combination of some of the above.

Project Parts

- Reaction Paper
  - Due March 18
- Project Proposal
  - Due March 25
- Final Report
  - Due May 1
- Presentation (in class, 12-min presentation + 3-min Q&A)
  - Scheduled for April 18, 23, 25

Project parts: Reaction Paper

- You get to pick two closely related papers to read
- Produce a reaction paper (2 to 3 pages) that contains:
  - Summary:
    - What is the main technical content of the papers?
    - How do the papers relate to the topics of the course?
    - What is the connection between the papers?
  - Critique:
    - What are the strengths and weaknesses of the papers?
    - Were there any unrealistic assumptions made?
  - Further work:
    - What are some promising further research questions along the directions of the papers?
    - How can the works be extended? Do you see an idea for a better model? A better algorithm?
    - Another problem or data the methods can be applied to?
- One of the goals is to help you generate a project idea for your proposal. (If you already have an idea, then you may choose the papers you read so that they have some relationship with your idea.)

Project parts: Proposal

- Ideally an outgrowth of the further work you identified in the Reaction Paper, but it could also be detached from it.
- Can also be chosen from a list of ideas I will provide.
- Roughly 2 to 3 pages.
- It should accomplish the following
  - Describe what you intend to do
  - Describe the methods you plan to use
  - Describe the data you will use and discuss how you plan to obtain it
  - Discuss relevant background work
  - Discuss your tentative plan

Project parts: Final Report

- A report of about 8 to 15 pages
- Content should roughly map to
  - *Introduction/Motivation/Problem Definition*
    - (where you state what you are trying to solve/achieve and why it matters)
  - *Model/Algorithm/Method*
    - (where you give a detailed description of your work)
  - *Results and findings*
    - (where you interpret the results you obtain, discuss implications, make observations and draw conclusions)
  - *Related work*
    - (where you cite and briefly summarize other work related to yours.)
- Later in the semester, I will provide a more detailed guideline on how to write an effective final report.

Datasets (may also inspire project ideas)

- Stanford Large Network Dataset Collection
- Co-authorship and citation networks
- Internet topology
- Stack overflow
- Yelp data
- Peer to peer money lending dataset
- YouTube dataset
- Amazon product copurchasing data
- Wikipedia
  - Page-to-page link data Dbpedia Edit history
- Movie ratings
- Who trusts whom data at Trustlet
- Interesting websites (for analysis):
  - Delicious Digg reddit StumbleUpon The Hype Machine Twitter jyte.com prosper.com lendingclub.com mturk.com

- http://snap.stanford.edu/data/
- http://www-personal.umich.edu/~mejn/netdata/

Software tools

- igraph (main tool used in the course)

Other (potentially relevant) tools:

- NetworKit (https://networkit.iti.kit.edu)

- SNAP (C++, Windows, Linux)
- Pajek (large network analysis, Windows, Linux)
- Networkx
- Gephi (visualization software)
- GUESS (exploratory data analysis and visualization)
- Infovis cyberinfrastructure (Linux, Windows, MacOSX)

Weeks ahead

- Next week
  - Tue Mar 5: Similarity
  - Thu Mar 7: Community identification
- Week after that
  - Tue Mar 12 (Spring break)
  - Thu Mar 14 (Spring break)
- Weak after that
  - Mon Mar 18 (Reaction Paper due)
  - Tue Mar 19 (class: Graph similarity with (un)known node correspondence)
  - Thu Mar 21 (class: Signed networks)